



HAL
open science

The impact of centromeres on spatial genome architecture

Héloïse Muller, Ines Anna Drinnenberg, José Gil Jr.

► **To cite this version:**

Héloïse Muller, Ines Anna Drinnenberg, José Gil Jr.. The impact of centromeres on spatial genome architecture. Trends in Genetics, 2019, 35 (8), pp.565-578. 10.1016/j.tig.2019.05.003 . hal-02189346

HAL Id: hal-02189346

<https://hal.science/hal-02189346>

Submitted on 19 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The impact of centromeres on spatial genome architecture

Héloïse Muller^{1,2*}, José Gil Jr.^{1,2} and Ines Anna Drinnenberg^{1,2}

¹Institut Curie, PSL Research University, CNRS, UMR3664, F-75005 Paris, France

²Sorbonne Université, Institut Curie, CNRS, UMR3664, F-75005 Paris, France

*Corresponding author: Muller, H. (heloise.muller@curie.fr)

Keywords:

Centromere, chromosome architecture, nuclear compartments, microscopy, Hi-C

Abstract:

The development of new technologies and experimental techniques are enabling researchers to see what was once unable to be seen. For example, the centromere was first seen as the mediator between spindle fiber and chromosome during mitosis and meiosis. Although this continues to be its most prominent role, we now know that the centromere functions beyond cellular division with important roles in genome organization and transcriptional regulation. Here we aim to share the structures and functions of centromeres in different organisms beginning with the diversity of their DNA sequence anatomies. We zoom out to describe their position in the nucleus and ultimately detail the different ways they contribute to genome organization and regulation at the spatial level.

1 **Centromeres: Beyond chromosome segregation**

2 Eukaryotic genomes are not randomly organized. In fact, chromatin is organized into
3 multiple domains juxtaposed one after the other along the length of a chromosome and
4 positioned into particular layouts and within particular regions of the nuclear space [1,2].
5 Among these chromatin domains, centromeres are universal to all eukaryotic chromosomes.
6 First cytologically described in 1882 by Walther Flemming as the primary constrictions on
7 condensed chromosomes attaching to spindle fibers [3], centromeres are essential for cell
8 division by ensuring the equal partitioning of DNA into daughter cells.

9 Although their crucial function comes into play during mitosis and meiosis, centromeres
10 are not inert during interphase. In fact, their importance in the genomes' 3D organization and
11 regulation has only started to be recognized. In diverse eukaryotes, centromeres occupy
12 particular domains in interphase nuclei. This non-random organization of centromeres in turn
13 impacts chromatin-based processes such as transcription and replication. Here, we review
14 centromere composition, localization, and interaction patterns across different eukaryotes.
15 Simultaneously, we highlight conserved and divergent principles of linear and spatial
16 centromeric architectures and their relationship with genome organization and function. For
17 this, we survey observations from microscopy and more recent genome conformation studies
18 to provide a systematic overview on spatial centromere organization in diverse eukaryotes.
19 We also discuss the functional relevance of particular centromeric arrangements in interphase
20 chromatin and emphasize current limitations and alternate strategies of genomic approaches
21 in determining the spatial organization of centromeres.

22 **Centromeres are enigmas of many genome assemblies**

23 Despite their essential function in mediating and controlling chromosome segregation,
24 centromere architecture is remarkably diverse among different organisms. In fact, the length
25 of centromeres ranges from as small as 120 base pairs to up to several megabases of DNA [4].
26 With some recent exceptions, sequence information is only available for shorter centromeres
27 whereas assemblies of larger centromeres are limited due to their length and repetitive
28 sequence composition (Figure 1, Box1).

1 Centromere sequences of many yeasts and fungal species have been characterized
2 (Figure 1, Box1). Comparative genomic analyses show that centromere size and **synteny** (see
3 glossary) are mostly conserved within both the *Saccharomycetaceae* and *Candida* clades,
4 pointing toward a common origin in each of the two clades. Nevertheless, centromere
5 sequences have rapidly diverged with very little or no detectable homology found between
6 species [5–10]. In contrast to the budding yeasts, the well-studied fission yeast
7 *Schizosaccharomyces pombe* has an incomplete genome assembly for two out of its three
8 centromeres with the number of pericentromeric inverted repeats still unknown (Figure 1,
9 Box 1) (www.pombase.org). Compared to fungi, sequence assemblies of the megabase-long
10 animal and plant centromeres, composed of satellite DNA interspersed with transposable
11 elements, are even more limited and remain largely unresolved even after using long-read
12 sequencing approaches [11–13]. Still, in some plant organisms, including *Oryzae sativa* and
13 *Arabidopsis thaliana*, the molecular structure of large centromeric regions for several
14 chromosomes has been determined (Figure 1, Box 1) [14–18]. In addition, a recent study
15 characterized the organization and sequence composition of all native *Drosophila*
16 centromeres, confirming the structure unraveled by original mapping efforts on the Dp1230
17 minichromosome (Figure 1, Box 1) [19,20]. Finally, the complete sequence of the first human
18 centromere has also been released by assembling the higher-order-repeat (HOR) structure of
19 the human chromosome Y satellite array (Figure 1, Box 1) [21].

20 Note that some species, including chicken, horse and potato, have repeat-based
21 centromeres as well as repeat-free centromeres depending on the chromosome [22–24]. For
22 instance, while most centromeres from the chicken macrochromosomes are composed of
23 chromosome-specific homogenous arrays of tandem repeats, chromosomes 5, 27 and Z are
24 composed of unique sequences. Comparative genomic studies show that the latter are likely
25 to represent evolutionarily more recent and “immature” centromeres that could represent
26 an intermediate state [25]. Consistent with this, **neocentromeres** (see glossary) often form on
27 sequences devoid of repetitive DNA [25] and tandem repeats have only recently been
28 accumulating in some **evolutionary new centromeres** (see glossary). Regardless of their
29 evolutionary framework, the non-repetitive nature of these centromeres provides a unique
30 opportunity allowing their unambiguous assembly.

1 **Seeing is believing: Centromeres follow a distinct nuclear organization**

2 While centromere assemblies are often limited, the location of centromeres can be
3 visualized by the presence of the histone variant CenH3 (also called CENP-A in humans), which
4 almost universally marks centromeric chromatin in eukaryotes [26]. Microscopy studies
5 provided the first insights into centromere organization in interphase nuclei. In numerous
6 fungal species, including *S. cerevisiae* and *S. pombe*, centromeres cluster all along the cell
7 cycle, which are visible as one large focus in the vicinity of the Spindle Pole Body (SPB) and
8 opposite to the nucleolus (Figure 2) [27–30]. Nevertheless, this organization can also vary in
9 some fungi. In *C. neoformans*, centromeres do not cluster in premeiotic cells and are instead
10 separately positioned adjacent to the nuclear envelope (Figure 2). Only when cells progress
11 toward mitosis do the centromeres gradually coalesce to a single cluster [31].

12 Chromosome organization in most plants has classically been divided into two main
13 categories: Rab1 or non-Rab1. The Rab1 organization is characterized by 1) all centromeres
14 being restricted to one pole of the nucleus, opposite to the nucleolus and 2) the telomeres
15 being more dispersed in the opposite pole. This organization is observed in wheat, rye, barley
16 and oat root tip cells (Figure 2) [32]. In contrast, sorghum and rice root cells adopt a non-Rab1
17 conformation with centromeres and telomeres dispersed and intermingled in the nuclear area
18 (Figure 2) [32]. Some species like maize have an intermediate conformation between Rab1 and
19 non-Rab1 in nuclei from root tip cells [32]. In addition, later studies modulated this original
20 classification by describing that chromosome organization can vary between different cell
21 types. For example, most root nuclei of rice show non-Rab1 organization, as described above,
22 while in the xylem vessel cells centromeres and telomeres localize to opposite poles of the
23 nuclei, consistent with a Rab1 configuration [32,33].

24 Compared to Rab1 and non-Rab1 conformation, *A. thaliana* has a distinct centromere
25 organization with (peri-)centromeric heterochromatin organized as clearly distinguishable
26 **chromocenters** (see glossary) in the nucleus (Figure 2). These are not randomly positioned
27 inside the nucleus and are found near the periphery or close to the nucleolus. Moreover, the
28 number of chromocenters observed is often (in ~88% of observed nuclei) less than 10
29 (corresponding to $2n = 10$ chromosomes), demonstrating a propensity to associate together
30 [34].

1 Centromere clustering in interphase nuclei has also been observed in several animal
2 model organisms. Counting the number of CENH3^{CID} positive foci in interphase nuclei using
3 immunofluorescence experiments demonstrated that centromeres cluster in fly cells [35]. In
4 particular, S2 aneuploid cells (13 stable chromosomes) show 4-6 clusters (Figure 2) and third-
5 instar larval diploid (2n=8) hemocytes show 2-3 clusters, which are localized close to the
6 periphery of the nucleolus in both cell types.

7 Centromere clustering is also a common feature in human and mouse. Early studies on
8 mouse B and T lymphocytes describe clustering of centromeres and their colocalization with
9 chromocenters comprised of pericentromeric heterochromatin (Figure 2) [36,37]. More
10 recent analyses of non-dividing/quiescent (G0) lymphocyte cells from mouse and humans
11 reveal that centromeres cluster preferentially at the nuclear periphery, forming on average
12 13 or 9 signals per human (2n=46) or mouse (2n=40) nuclei, respectively [38]. In the few cases
13 of a central localization, clusters are found to be associated with the nucleolus. Refining
14 centromere locations within the nuclear region occupied by their corresponding chromosome
15 (also called **chromosome territory** (CT, see glossary), centromeres are found at the periphery
16 of CTs in 95% of studied cases and positioned towards the nuclear periphery. Other studies
17 confirm and expand these results showing that centromere clustering is a common feature
18 observed to various extents in all cell types analyzed [39–42], but changes dynamically over
19 the course of the cell cycle [41,42]. The cell types analyzed include hematopoietic progenitors
20 and mature blood cells (B and T lymphocytes, granulocytes, monocytes) as models for cell
21 differentiation, as well as fibroblasts, which offer a comparison with a terminally
22 differentiated cell from another lineage. Upon exit of mitosis, in early G1, very little to no
23 clustering is observed and the number of kinetochore signals decrease in late G1 and S phases
24 indicating centromere clustering. Over the course of G2, centromeres separate out again as
25 chromosomes start to condense in preparation for mitosis. In addition, peripheral localization
26 and clustering of centromeres are less pronounced in nuclei of cycling cells or progenitor cells
27 compared to non-cycling and terminally differentiated cells [41,42]. In particular, while
28 centromere clustering is maintained in human embryonic stem cells, their peripheral
29 localization is mainly lost, possibly due to rapid divisions of these cells [43].

1 In summary, while the insights gained from microscopy approaches are limited, these
2 studies reveal that centromeres are often confined to or even clustered in specific nuclear
3 locations in many organisms. Their organization within the nucleus, however, varies from one
4 species to another.

5 **Centromere interactions drive genome 3D organization and regulation**

6 Genome-wide chromosome architecture studies using high-throughput molecular
7 biology techniques, like **Chromosome Conformation Capture** (3C, see glossary) and its
8 derivatives (**4C**, **Hi-C**,...; see glossary) offer new ways to study the importance of centromeres
9 in genome organization. Early studies define the top of genome organization hierarchy as
10 being CTs [44]. In line with that, Hi-C data from different organisms describe that
11 intrachromosomal interactions are always more frequent than interchromosomal
12 interactions. In many organisms, and especially in mammals, chromosomes can be further
13 segregated into two compartments: the A compartment, which includes regions of the
14 genome enriched in genes and histone modifications associated with active chromatin, or the
15 B compartment, which is enriched in histone modifications associated with inactive chromatin
16 [45]. Going down the chromosome organization to the scale of several hundred kilobases, one
17 can find domains which interact more frequently within themselves than with neighboring
18 regions, often referred to as topologically associated domains (TADs) [46]. In addition, other
19 signatures of genome organization are also visible on Hi-C maps, and account for specific and
20 strong landmarks of 3D genome organization. Here, the centromere is a classic example of a
21 region with strong topological impact. Assembling data from many organisms published to
22 date, we will discuss at least two implications to previously described centromere interactions
23 visible on Hi-C matrices: (i) centromeres build a specific sub-compartment inside the nucleus
24 and (ii) they form a barrier to intrachromosomal arm interactions.

25 *Centromeres form strong interacting sub-compartments*

26 Three-dimensional genome organization has been studied in several fungal species. The
27 first *S. cerevisiae* full genome **contact map** (see glossary) was published using a high
28 throughput 4C-derived experiment [47]. From this map, centromere clustering was reported
29 to be the most striking feature of interchromosomal contacts, in accordance with the known

1 Rabl-like organization of budding yeast chromosomes. Aside from that, centromeres engage
2 in relatively few long-range intrachromosomal interactions (Figure 3A). These findings have
3 been confirmed in later studies of other budding yeasts, filamentous fungi and *S. pombe* [48–
4 52]. Centromere clustering is such a prominent characteristic of yeast genome organization
5 that the 3C signal has even been taken as a starting point to locate previously unidentified
6 centromeres [53,54].

7 The remarkable plasticity of yeast genomes has allowed to drastically remodel the
8 location and the number of centromeres. For example, selecting for neocentromeres in *C.*
9 *albicans*, or manipulating the chromosome number by genome engineering in *S. cerevisiae*,
10 result in a drastic reorganization of genomic contacts [55–57]. In both cases, upon the loss of
11 a centromere, the strong interactions of flanking sequences with other centromeric regions
12 disappear. In *C. albicans*, chromosomal regions near a neocentromere acquire new and strong
13 interactions with all the other seven centromeric regions. The impact of this reorganization
14 on gene expression was also analyzed because co-regulated genes often cluster along
15 chromosome arms and in centromere-proximal regions [58,59]. In this case however, the
16 impact on gene expression upon centromere delocalization is marginal, at least under
17 standard laboratory growth conditions. In addition, reshuffling centromere location might also
18 interfere with DNA metabolism and recombination. In fact, centromeres are among the
19 earliest replicated regions in the yeast genome [48,60,61] and negatively impact the rate of
20 cross-over during meiosis [62]. Future studies could therefore investigate the effects of
21 centromere delocalization on other chromatin processes apart from gene expression.

22 Centromeric and pericentromeric regions also drive specific physical contacts in
23 interphase of other organisms as visualized on their respective Hi-C contact maps (Figure 3).
24 In line with their Rabl or non-Rabl organization, two different types of contact maps have been
25 described in plants. The contact map of barley shows a strong cross shape signal within and
26 between chromosomes, consistent with their typical Rabl conformation [63]. In contrast, the
27 contact map of rice, tomato, foxtail millet and sorghum, that are described as non-Rabl, show
28 a tripartite segregation of their chromosomes with centromeres and large pericentromeric
29 heterochromatin grouped in a central B (inactive) compartment. Although interactions are
30 enriched within all inactive compartments across the genome, they tend to be less involved

1 in interchromosomal interactions than the two distal A (active) compartments [32,64]. Though
2 maize was previously reported to have an intermediate organization, Hi-C analysis
3 demonstrated a strong intra- and interchromosomal cross signal, indicating a similar Rab1
4 organization as in barley, at least in mesophyll protoplasts cells [32,64]. The genome-wide
5 contact map of *A. thaliana* shows detectable interactions between all pairs of centromeric
6 regions that make very little contacts with other genomic regions, except for the large
7 heterochromatic locus on chromosome 4 called “knob” (Figure 3B) [65]. Here, centromere
8 clustering gives rise to a less pronounced cross shape signal than observed in plants with Rab1
9 conformation. Still, the strongest genome-wide Hi-C contacts of *A. thaliana* correspond to
10 pericentromeric chromatin interactions, both within the same pericentromere and between
11 different pericentromeres, consistent with the reported formation of heterochromatic
12 chromocenters [65].

13 Centromere-centromere contacts are also visible on the *D. melanogaster* Hi-C matrix
14 which shows strong interaction signals between the megabase long, heterochromatic, peri-
15 centromeric regions (Figure 3C) [66]. To explain these interactions, as well as other general
16 chromosomal organization in flies (see below), it is hypothesized that (peri-)centromeric
17 regions may assemble into their own nuclear compartment, again consistent with previous
18 observations by microscopy [67].

19 Genome-wide Hi-C maps in vertebrates also show the presence of CTs, but centromere
20 interactions are not as apparent (Figure 3D, E) [45,68]. The lack of strong centromere signals
21 in vertebrate Hi-C contact maps is possibly due to weaker centromere interactions in these
22 organisms. In addition, the lack of full centromere sequence assemblies might also be a
23 confounding factor. Satellite sequences are systematically excluded from most next
24 generation sequencing analysis due to their repetitive nature. Nevertheless, as the interaction
25 signal spreads along neighboring chromatin fibers due to their physical properties, it is
26 possible to recover the interaction encompassing repetitive regions by looking at adjacent
27 non-ambiguous regions. Using this principle to reanalyze human and mouse Hi-C data, it was
28 found that decomposing the contact map signal into a series of genomic tracks, known as
29 eigenvectors, enable to determine the contribution of particular genomic features to higher-
30 order chromosomal organization [69]. They further determine that the first eigenvector

1 relates to genomic sequences and local epigenetic chromatin states (as described before by
2 the compartmentalization), while the second (or third in the case of mouse) relates to the
3 position along the chromosome arm and highlights centromere-centromere interactions. The
4 association of all repetitive elements in the human and mouse genomes has also been studied
5 by using neighboring regions of ambiguous sequences, as well as stringent and restrictive
6 mapping parameters [70]. With these techniques, it was shown that some satellites
7 significantly co-localize inside the nucleus, particularly alpha-satellites in the human genome.
8 Furthermore, by using higher resolution maps of the human genome (down to 1 kb bins), the
9 two original compartments that were described by the first eigenvector were refined and
10 divided into five sub-compartments, two in A and three in B [71]. Sixty-two percent of
11 pericentromeric heterochromatin is found in sub-compartment B2, which is associated with
12 the nuclear lamina and nucleolus associated domains. Finally, superimposition of 3D genome
13 structure determination from single-cell Hi-C data with microscopy visualizing CENP-A reveal
14 a Rab1-conformation with centromeres and telomeres clustered at opposite sides of the
15 nucleus in G1 mouse ES cells [72]. Taken together, centromere clustering in vertebrates,
16 although less apparent in Hi-C contact maps, still account for a prevalent class of contacts that
17 have been demonstrated by many different means.

18 To overcome the problem of mapping repetitive sequences using a different strategy,
19 centromere interactions were analyzed in chicken cells, taking advantage of the non-repetitive
20 nature of centromeres on chromosomes 5, 27 and Z, as well as engineered strains with
21 neocentromeres at several positions on the Z chromosome [73]. As in humans,
22 neocentromeres in chicken are neither flanked by heterochromatin nor enriched for satellite
23 repeat sequences [74–76]. Using this set-up, it has been shown that repetitive and non-
24 repetitive centromeres associate with one another in 4C analyses and that non-repetitive
25 centromeres and neocentromeres interact with heterochromatic regions of the genome.

26 Centromere clustering even extends to Apicomplexans, single-cell organisms that
27 include the human malaria parasite *Plasmodium falciparum*. Recent Hi-C analyses of several
28 *Plasmodium* species and two additional apicomplexan parasites reveal varying degrees of
29 centromere and telomere clustering in most developmental stages. However, centromere

1 interactions appear to be lost in sporozoites, haploid cells that invade the insect salivary glands
2 for transmission to the host [77,78].

3 To conclude, it is important to note that while the described features often arise from
4 asynchronous cell populations they are unlikely the result of interactions coming from the low
5 fraction of mitotic cells, as previously proposed in studies using synchronized cells [79–81].
6 Mitotic chromosomes are devoid of compartments and TADs and enriched in
7 intrachromosomal long-range contacts compared to interphase chromosomes. This particular
8 folding makes them drastically different from interphase chromosomes that resemble and
9 better recapitulate chromosome conformation in exponentially growing cells. Taken together,
10 despite different linear genome architectures and centromere sequence compositions,
11 centromere interactions appear to be a common feature in many organisms scattered
12 throughout the eukaryotic phylogeny.

13 *Centromeres individualize chromosome arms*

14 In addition to engaging in preferential interactions and building a specific sub-
15 compartment, centromeres tend to create a barrier within each chromosome resulting in less
16 frequent contacts between the chromosome arms on either side of them, compared to intra-
17 arm contact frequencies (Figure 3F). This insulation property is visible on contact maps of
18 species harboring a classical Rab1 or Rab1-like chromosome conformation (Figure 3).
19 Measurements in *S. pombe* for instance show more frequent intra-arm than inter-arm
20 interactions; the latter being in the same range as interaction frequencies between
21 chromosomes [50]. The barrier effect of centromeres also plays a role in species with other
22 genome configurations. The first 3D studies in *A. thaliana* applying 4C on several viewpoints
23 distributed across all five chromosomes demonstrate that chromosome arms are the main
24 interaction units of the genome [82]. This result was later confirmed on genome-wide analysis
25 showing a higher level of interactions between chromatin regions on the same side of a
26 centromere, as opposed to interactions on opposite sides [65].

27 Genomic rearrangements involving the centromere in mutant flies demonstrate the
28 insulation properties of centromeres. Interactions between polycomb target genes in *D.*
29 *melanogaster* wildtype and in a mutant strain carrying a **pericentromeric inversion** (see

1 glossary) on chromosome 3 has been analyzed using 4C. This demonstrated preferential
2 interactions between these loci when present on the same chromosome arm. These findings
3 imply a strong rewiring of chromatin interactions in the rearranged strain consistent with a
4 model of distinct territories formed by individual chromosome arms [67]. This model,
5 however, has to be modulated depending on the developmental stage or tissue analyzed
6 because relative interactions between arms and chromosomes are reported to vary
7 [66,83,84].

8 Analyses in human cells using a modified 3C technique, called Tethered Conformation
9 Capture (TCC), reveal that the contact profile of inactive regions decreases abruptly when
10 located on opposite sides of the centromere [85]. This effect, however, is not seen for active
11 regions which tend to be involved in long-range contacts regardless of the presence of the
12 centromere [85]. These analyses demonstrate that the human centromere can also act as a
13 contact barrier, at least for some categories of compartments.

14 In addition, the position of the centromere on the chromosome also constrains its
15 overall architecture. This effect can be observed in *S. cerevisiae* where the small and long
16 chromosome arms tend to interact more frequently with one another, respectively [86,87].
17 Furthermore, the position of the centromere on the chromosome also influences centromere-
18 centromere interactions themselves. For instance, centromeric regions of human acrocentric
19 chromosomes are more likely to contact each other than those of metacentric chromosomes
20 [85].

21 Overall, the centromere influences interaction frequencies along the entire
22 chromosome preventing certain contacts between particular chromosomal regions. Variation
23 in the position of the centromere can directly impact long distance interactions and the
24 general 3D conformation of chromosomes. It will be important to determine to what extent
25 altered centromere locations can lead to perturbations in gene expression or other DNA
26 related metabolisms in future studies.

27 **Concluding remarks and future perspectives**

28 In this review, we provide a collective view on the spatial organization of centromeres
29 in a range of distinct eukaryotes (Key table). Despite their essential function, centromere

1 architectures and sequences are highly diverse. Yet, in many organisms the spatial
2 organization of centromeres appears to be conserved. Centromeres localize to distinct nuclear
3 sub-compartments or even cluster with one another in interphase nuclei. Given that the first
4 layer of genome organization is the segregation of chromosomes into CTs with
5 intrachromosomal interactions being the most frequent of all, centromere clustering across
6 different chromosomes is not an intuitive result. This raises the questions of what leads to
7 centromere clustering and whether there is a functional relevance to this organization (see
8 outstanding question box).

9 Centromere clustering could be mediated by proteins that bind to centromeres or
10 pericentromeres and stabilize long-range interactions between them. This mechanism could
11 be similar to the formation of chromocenters in flies and mice that were proposed to be
12 mediated by proteins bound to pericentromeric satellites and capable of bundling multiple
13 DNA strands [88]. Future studies could aim to identify additional proteins involved in the
14 formation of centromere clusters. The physical properties of centromere clusters could phase
15 separate them into nuclear sub-compartments as it has been proposed to underlie A/B
16 compartmentalization observed in Hi-C data [89] and the formation of heterochromatin foci
17 bound by HP1 [90]. In fact, kinetochore components are enriched in coiled-coil and disordered
18 domains [91] which represent structural features frequently associated with protein
19 assemblies that induce protein phase separation [92].

20 Several observations point towards a functional relevance of centromere clustering.
21 Recent studies have hypothesized that the aggregation of chromosomes via chromocenters in
22 interphase could actually prevent micronuclei formation and thus play a role in genome
23 stability [88]. This is consistent with observations showing that destabilization of spatial
24 centromere positioning impairs silencing which in turn leads to an increase in transcription of
25 transposable elements. This is associated with the accumulation of DNA double-strand breaks
26 throughout the genome leading to mitotic defects and genome instability [35,93,94].
27 Additionally, centromere organization is subject to changes during the cell cycle and during
28 various developmental stages including plant germination or floraison and mouse
29 embryogenesis [95,96]. Several studies have also pointed out possible higher order cluster
30 organization in human cells, with specific combinations of centromere associations

1 [39,40,85,97,98]. Hence, the question arises whether such non-random centromere
2 associations have functional implications on surrounding chromatin processes. Still, given the
3 dominant role of the cell cycle in determining genome architecture, temporal changes in cell-
4 cycle stages will have to be taken into consideration when comparing centromere
5 configurations across different organisms and developmental stages.

6 The position of the centromere on the chromosome together with its 3D organization
7 also impacts the chromatin environment in its vicinity. It will therefore be interesting in future
8 studies to investigate how genome architecture changes with centromere organization. This
9 aspect could be addressed in organisms that have evolved neocentromeres or contain
10 evolutionary new centromeres by comparing their genetic environment to the ancestral state.
11 Additionally, studies in organisms with drastically different centromere organization, such as
12 **holocentric** (see glossary) species, could also provide insights into the impact of centromeres
13 on genome architecture.

14 Finally, a comprehensive picture on the conservation of 3D centromere conformation
15 will require data from additional organisms including non-model organisms. However, in order
16 to do this, we will have to overcome the challenges brought forth by the repetitive nature of
17 many centromeres that result in incomplete genome assemblies. New and improved
18 sequencing technologies will be required to get insights into centromere configurations in
19 organisms with repetitive centromeres including humans.

20

21

1 **Box1: Centromere organization is diverse among eukaryotes.**

2 *S. cerevisiae* “point” centromeres are genetically defined by three genetic elements:
3 CDEI, II and III and incorporate a single centromeric specific histone H3 variant, CenH3 (Cse4)
4 [99,100].

5 In the *Candida* clade, including *C. albicans*, centromeres occur in large ORF-free regions
6 of 4-18 kb which includes a CenH3-containing-nucleosome core region of 3-5 kb. They share
7 no common sequence motif or repeats either within one organism or between them [5–7].

8 In *S. pombe*, the central core (cnt), of about 4 kb in length, assembles CenH3-containing-
9 nucleosomes. It is surrounded by numerous alternating dg, dh and cen253 elements forming
10 innermost (imr) and outermost (otr) repeats, giving rise to centromeric regions of 35–120 kb
11 [101]. A similar organization with a non-repetitive mid core (2-5 kb) flanked by inverted
12 repeats (2-5 kb) is found in the phylogenetically distant yeast *Candida tropicalis* [102].

13 In *Cryptococcus* species, centromeres are in syntenic positions in ORF-free, poorly
14 transcribed regions featuring variable combinations of retrotransposons or their remnants.
15 Centromeres span from 44 and 62 kb on average in *C. neoformans* and *C. deneoformans*,
16 respectively, and about 14 kb in *C. deuterogatii* [103].

17 *A. thaliana* contains centromeres of several megabases with a central domain consisting
18 mainly of 180 bp satellite DNA and fragments of Athila retrotransposon elements, 106B. The
19 flanking regions contain multiple families of LTR retrotransposons and 5SrDNA sequences
20 [15,16,104].

21 Rice and maize contain satDNA, named CentO (155 bp) and CentC (156 bp), respectively.
22 Each also contains centromeric-specific retrotransposons, CRM (-Maize) and CRR (-Rice); while
23 in wheat, CRW (-Wheat) are found, but no satDNA [14,17,18,105–108].

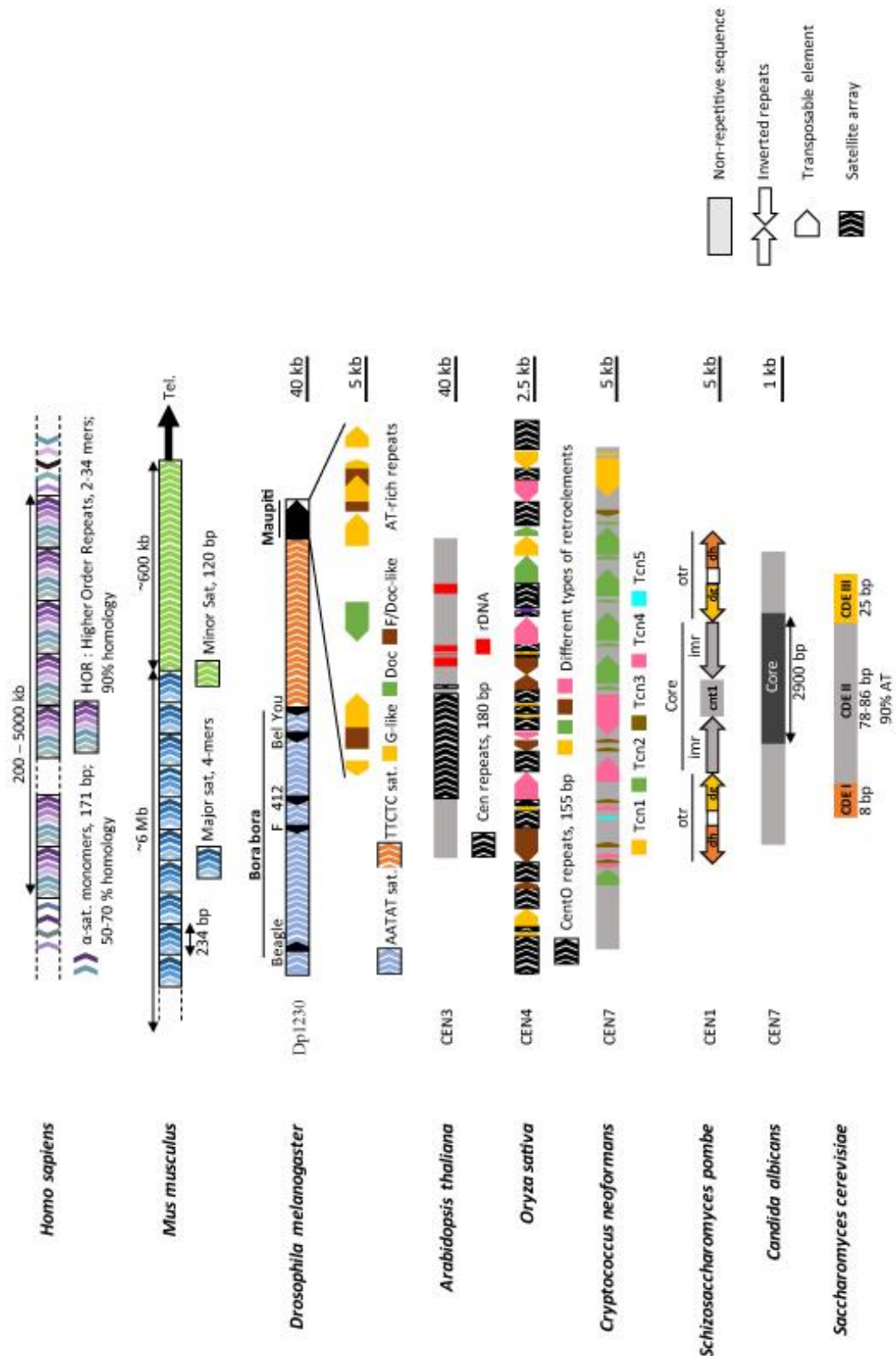
24 Analysis of the *D. melanogaster* centromere organization revealed that they are
25 composed of islands of complex DNA enriched in retrotransposons flanked by arrays of short
26 satellite repeats (5 to 12 bp) [19,20]. Whether satellite or retrotransposon sequences are the
27 functional parts of *D. melanogaster* centromeres based on CENP-A enrichment remains
28 however controversial [109].

1 Mouse chromosomes are all acrocentric, with centromeres composed of 120 bp minor-
2 satellites constituting the core centromeric region toward the telomere and the more
3 abundant 234 bp major-satellite forming pericentromeric heterochromatin [110,111].

4 The centromere of the human Y chromosome is made up of 171 bp alpha-satellite
5 monomers organized in a tandem fashion into large arrays of HOR [21]. While annotated
6 alpha-satellite arrays are dominated by HORs on BAC clones conferring chromosome-
7 specificity [112], another study proposes that the core of human centromeres is composed of
8 two classes of highly homogenous alpha-satellite dimers with HORs being pericentromeric
9 [113].

10

11



2 **Figure 1. Sequence organization of centromeres in eukaryotes.** For each species, the
 3 typical centromere organization is shown, with the size range for features of interest. When
 4 the centromere of a particular chromosome is depicted, its number is indicated after the
 5 species name and an approximate scale is shown on the right.

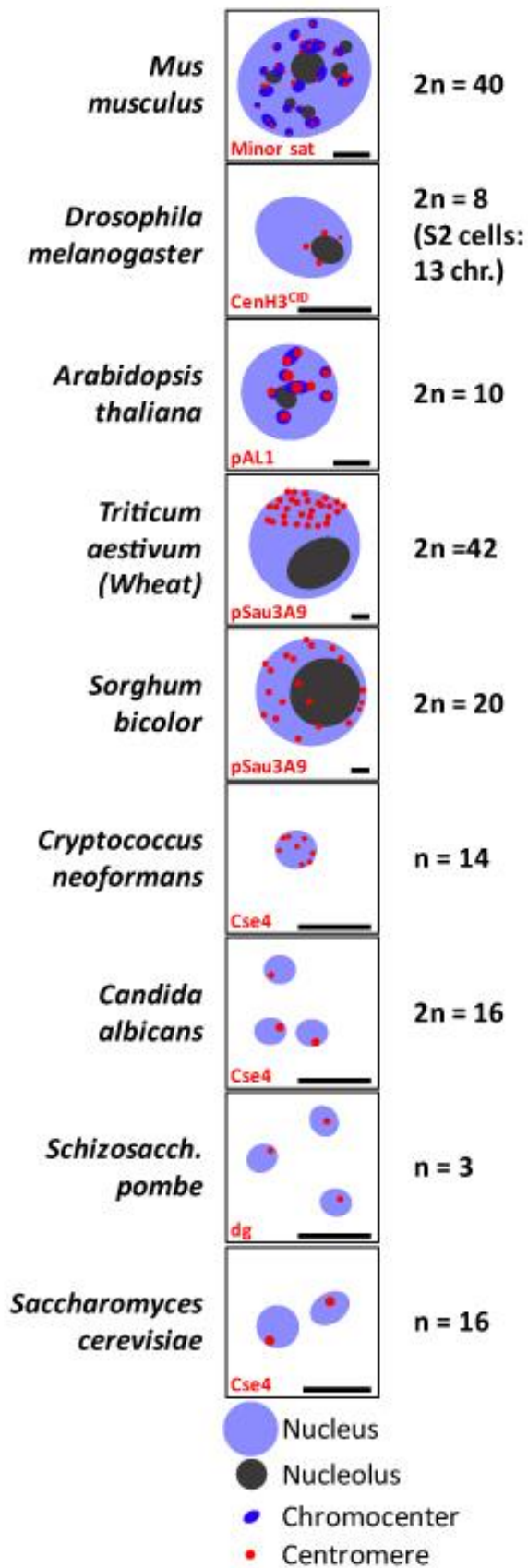
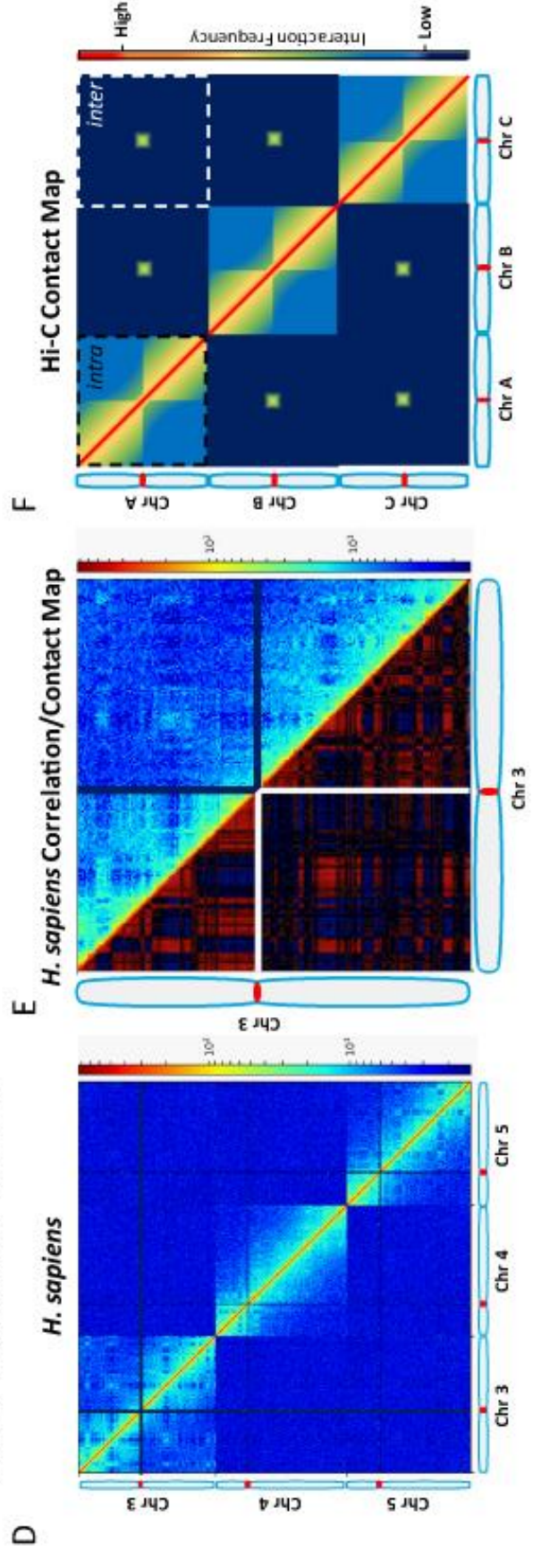
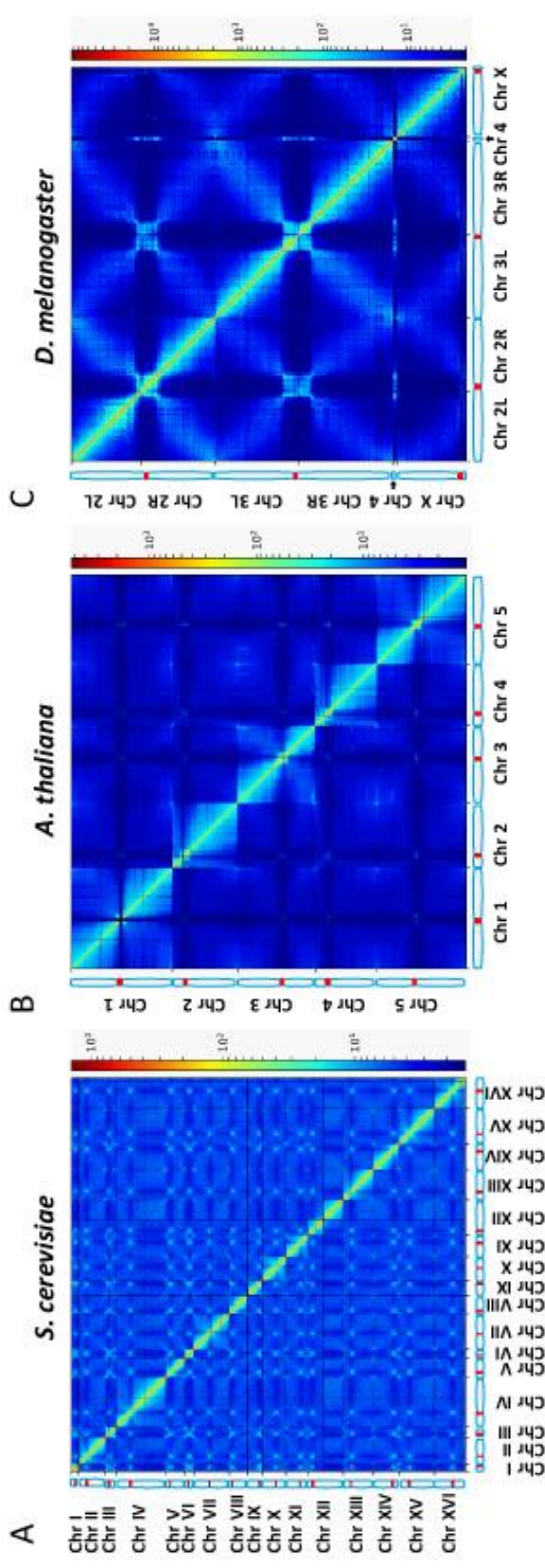


Figure 2. Schematic of centromere localization in interphase. Blue: DAPI, red: centromere probe by IF or FISH as indicated. Adapted from: *M. musculus*: [114]; *D. melanogaster*: [35]; *A. thaliana*: [34]; *T. aestivum* and *S. bicolor*: [32]; *C. neoformans*: [31]; *C. albicans*: [30]; *S. pombe*: [27]; *S. cerevisiae*: [115]. Scale bar = 5 μ m.



1 **Figure 3. Normalized contact maps** of (A) *S. cerevisiae*, full genome, resolution 5 kb,
2 dataset from [86] (B) *A. thaliana*, full genome, resolution 20 kb, dataset from [65] (C) *D.*
3 *melanogaster*, full genome, resolution 100 kb, dataset from [84]. (D) GM06990 human cell
4 line, chr 3-5, resolution 1 Mb, dataset from [45]. (E) Chr 3 as in D; top half: normalized contact
5 map; bottom half: **correlation matrix** (see glossary). Contact maps have been generated using
6 the HiC-Pro suite, normalized by ICE using program default parameters [116] and drawn using
7 HiCExplorer [117]. Correlation matrix has been generated by HiTC R package [118]. (F)
8 Schematic of a simplified Hi-C contact map displaying three chromosomes. In the
9 intrachromosomal area (black dashed box), chromosome arms are individualized on each side
10 of the centromere. Centromere-centromere interactions are the main features observed in
11 the interchromosomal area (white dashed box).

12

1 **Key table: Centromere organization in eukaryotes.**

Species	Nb of chr.	Type of centromere	Centromere imaging	Chromosome organization
<i>S. cerevisiae</i>	n = 16 2n = 32	Point centromeres	One focus near SPB	Rabl-like
<i>C. albicans</i>	2n = 16	Regional without repetitive sequence	One focus near SPB	Rabl-like
<i>S. pombe</i>	n = 3 2n = 6	Regional with inverted repeats	One focus near SPB	Rabl-like
<i>C. neoformans</i>	n = 14	Regional with transposable elements	Individual foci at nuclear periphery	ND
<i>A. thaliana</i>	2n = 10	Regional with transposable elements and SatDNA	Clustered foci at chromocenters	Centromere clustering
<i>O. sativa</i> (rice) / <i>S. bicolor</i>	2n = 24 / 2n = 20	Regional with transposable elements and SatDNA	Dispersed foci in nucleus	Non-Rabl
<i>T. aestivum</i> (wheat)	2n = 6X =42	ND	Clustered foci at one cell pole, opposite to nucleolus	Rabl
<i>D. melanogaster</i>	2n = 8	Regional with transposable elements and SatDNA	Clustered foci around the nucleolus	Chromosome territory and arm territory
<i>M. musculus</i>	2n = 40	Regional with SatDNA	Clustered foci at chromocenters localized at nuclear periphery	Chromosome territory
<i>H. sapiens</i>	2n = 46	Regional with SatDNA	Clustering at nuclear periphery	Chromosome territory

2 ND: not determined.

3

1 Glossary:

2 **Synten**y: conservation of the order of homologous genetic loci between two
3 chromosomes.

4 **Neocentromeres**: new centromeres that form at an ectopic locus on the chromosome
5 that is not centromeric. They typically arise due to the disruption of the endogenous
6 centromere. Neocentromeres are often not associated with repetitive DNA but instead consist
7 of unique sequences.

8 **Evolutionary new centromeres**: recently evolved novel centromeres appearing at
9 ectopic locations on the chromosome compared to an ancestral centromere. Like
10 neocentromeres, evolutionary new centromeres are often devoid of repetitive DNA.

11 **Chromocenter**: densely staining aggregation of heterochromatic regions in the nucleus
12 of some cells.

13 **Chromosome Territory**: discrete region of the cell nucleus preferentially occupied by a
14 particular chromosome.

15 **Chromosome Conformation Capture (3C)**: method in molecular biology to analyze the
16 spatial organization of chromatin in a cell between two specified loci.

17 **4C**: variation of 3C to visualize the contact pattern of one locus to any loci genome-wide.

18 **Hi-C**: variation of 3C in which biotin is incorporated at the ligation junction enabling it to
19 be selected and obtain genome-wide contact maps.

20 **Contact map**: representation of a Hi-C experiment as a diagonally symmetrical square
21 matrix indicating the frequency of contacts between any pairs of genomic loci.

22 **Pericentromeric Inversion**: an inversion is a chromosome rearrangement in which a
23 segment of a chromosome is reversed end to end. In a pericentric inversions the segment
24 includes the centromere.

25 **Holocentric**: referring to chromosomes that have multiple centromeres along their
26 entire length, as opposed to monocentric where centromeres are restricted to one region on
27 each chromosome.

28 **Correlation matrix**: normalized contact map where the Pearson correlation is taken
29 between the i^{th} row and j^{th} column, which sharpens the plaid pattern.

30

31

1 References:

- 2 1 Rowley, M.J. and Corces, V.G. (2018) Organizational principles of 3D genome architecture. *Nat.*
3 *Rev. Genet.* 19, 789–800
- 4 2 Towbin, B.D. *et al.* (2013) Mechanisms of heterochromatin subnuclear localization. *Trends*
5 *Biochem. Sci.* 38, 356–363
- 6 3 Flemming, W. (1882) *Zellsubstanz, Kern Und Zelltheilung.*, F.C.W. Vogel.
- 7 4 Steiner, F.A. and Henikoff, S. (2015) Diversity in the organization of centromeric chromatin. *Curr.*
8 *Opin. Genet. Dev.* 31, 28–35
- 9 5 Sanyal, K. *et al.* (2004) Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are
10 all different and unique. *Proc. Natl. Acad. Sci.* 101, 11374–11379
- 11 6 Mishra, P.K. *et al.* (2007) Centromere size and position in *Candida albicans* are evolutionarily
12 conserved independent of DNA sequence heterogeneity. *Mol. Genet. Genomics MGG* 278, 455–
13 465
- 14 7 Padmanabhan, S. *et al.* (2008) Rapid evolution of Cse4p-rich centromeric DNA sequences in
15 closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. *Proc. Natl. Acad.*
16 *Sci. U. S. A.* 105, 19797–19802
- 17 8 Dujon, B. *et al.* (2004) Genome evolution in yeasts. *Nature* 430, 35–44
- 18 9 Gordon, J.L. *et al.* (2011) Mechanisms of chromosome number evolution in yeast. *PLoS Genet.* 7,
19 e1002190
- 20 10 Meraldi, P. *et al.* (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore
21 proteins. *Genome Biol.* 7, R23
- 22 11 Chin, C.-S. *et al.* (2016) Phased diploid genome assembly with single-molecule real-time
23 sequencing. *Nat. Methods* 13, 1050–1054
- 24 12 Michael, T.P. *et al.* (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single
25 nanopore flow cell. *Nat. Commun.* 9, 541
- 26 13 Solares, E.A. *et al.* (2018) Rapid Low-Cost Assembly of the *Drosophila melanogaster* Reference
27 Genome Using Low-Coverage, Long-Read Sequencing. *G3amp58 GenesGenomesGenetics* 8,
28 3143–3154
- 29 14 Cheng, Z. *et al.* (2002) Functional rice centromeres are marked by a satellite repeat and a
30 centromere-specific retrotransposon. *Plant Cell* 14, 1691–1704
- 31 15 Kumekawa, N. *et al.* (2000) The size and sequence organization of the centromeric region of
32 *Arabidopsis thaliana* chromosome 5. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 7, 315–321
- 33 16 Kumekawa, N. *et al.* (2001) The size and sequence organization of the centromeric region of
34 *Arabidopsis thaliana* chromosome 4. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 8, 285–290
- 35 17 Wu, J. *et al.* (2004) Composition and structure of the centromeric region of rice chromosome 8.
36 *Plant Cell* 16, 967–976
- 37 18 Zhang, Y. *et al.* (2004) Structural features of the rice chromosome 4 centromere. *Nucleic Acids*
38 *Res.* 32, 2023–2030
- 39 19 Chang, C.-H. *et al.* (2019) Islands of retroelements are the major components of *Drosophila*
40 centromeres. *bioRxiv* DOI: 10.1101/537357
- 41 20 Sun, X. *et al.* (2003) Sequence analysis of a functional *Drosophila* centromere. *Genome Res.* 13,
42 182–194
- 43 21 Jain, M. *et al.* (2018) Linear assembly of a human centromere on the Y chromosome. *Nat.*
44 *Biotechnol.* 36, 321–323
- 45 22 Gong, Z. *et al.* (2012) Repeatless and repeat-based centromeres in potato: implications for
46 centromere evolution. *Plant Cell* 24, 3559–3574
- 47 23 Piras, F.M. *et al.* (2010) Uncoupling of satellite DNA and centromeric function in the genus *Equus*.
48 *PLoS Genet.* 6, e1000845

1 24 Shang, W.-H. *et al.* (2010) Chickens possess centromeres with both extended tandem repeats
2 and short non-tandem-repetitive sequences. *Genome Res.* 20, 1219–1228

3 25 Nergadze, S.G. *et al.* (2018) Birth, evolution, and transmission of satellite-free mammalian
4 centromeric domains. *Genome Res.* 28, 789–799

5 26 Musacchio, A. and Desai, A. (2017) A Molecular View of Kinetochores Assembly and Function.
6 *Biology* 6,

7 27 Funabiki, H. *et al.* (1993) Cell cycle-dependent specific positioning and clustering of centromeres
8 and telomeres in fission yeast. *J. Cell Biol.* 121, 961–976

9 28 Guacci, V. *et al.* (1997) Centromere position in budding yeast: evidence for anaphase A. *Mol. Biol.*
10 *Cell* 8, 957–972

11 29 Jin, Q.W. *et al.* (2000) Centromere clustering is a major determinant of yeast interphase nuclear
12 organization. *J. Cell Sci.* 113 (Pt 11), 1903–1912

13 30 Thakur, J. and Sanyal, K. (2012) A coordinated interdependent protein circuitry stabilizes the
14 kinetochores ensemble to protect CENP-A in the human pathogenic yeast *Candida albicans*. *PLoS*
15 *Genet.* 8, e1002661

16 31 Kozubowski, L. *et al.* (2013) Ordered Kinetochores Assembly in the Human-Pathogenic
17 Basidiomycetous Yeast *Cryptococcus neoformans*. *mBio* 4, e00614-13-e00614-13

18 32 Dong, F. and Jiang, J. (1998) Non-Rabl patterns of centromere and telomere distribution in the
19 interphase nuclei of plant cells. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome*
20 *Biol.* 6, 551–558

21 33 Prieto, P. *et al.* (2004) Chromosomes associate premeiotically and in xylem vessel cells via their
22 telomeres and centromeres in diploid rice (*Oryza sativa*). *Chromosoma* 112, 300–307

23 34 Franz, P. *et al.* (2002) Interphase chromosomes in *Arabidopsis* are organized as well defined
24 chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14584–
25 14589

26 35 Padeken, J. *et al.* (2013) The nucleoplasmin homolog NLP mediates centromere clustering and
27 anchoring to the nucleolus. *Mol. Cell* 50, 236–249

28 36 Brown, K.E. *et al.* (1997) Association of transcriptionally silent genes with Ikaros complexes at
29 centromeric heterochromatin. *Cell* 91, 845–854

30 37 Vourc'h, C. *et al.* (1993) Cell cycle-dependent distribution of telomeres, centromeres, and
31 chromosome-specific subsatellite domains in the interphase nucleus of mouse lymphocytes. *Exp.*
32 *Cell Res.* 205, 142–151

33 38 Weierich, C. *et al.* (2003) Three-dimensional arrangements of centromeres and telomeres in
34 nuclei of human and murine lymphocytes. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp.*
35 *Chromosome Biol.* 11, 485–502

36 39 Alcobia, I. *et al.* (2003) The spatial organization of centromeric heterochromatin during normal
37 human lymphopoiesis: evidence for ontogenically determined spatial patterns. *Exp. Cell Res.* 290,
38 358–369

39 40 Alcobia, I. *et al.* (2000) Spatial associations of centromeres in the nuclei of hematopoietic cells:
40 evidence for cell-type-specific organizational patterns. *Blood* 95, 1608–1615

41 41 Bartholdi, M.F. (1991) Nuclear distribution of centromeres during the cell cycle of human diploid
42 fibroblasts. *J. Cell Sci.* 99 (Pt 2), 255–263

43 42 Solovei, I. *et al.* (2004) Differences in centromere positioning of cycling and postmitotic human
44 cell types. *Chromosoma* 112, 410–423

45 43 Wiblin, A.E. (2005) Distinctive nuclear organization of centromeres and regions involved in
46 pluripotency in human embryonic stem cells. *J. Cell Sci.* 118, 3861–3868

47 44 Cremer, T. and Cremer, M. (2010) Chromosome territories. *Cold Spring Harb. Perspect. Biol.* 2,
48 a003889

49 45 Lieberman-Aiden, E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals
50 folding principles of the human genome. *Science* 326, 289–293

1 46 Dekker, J. and Mirny, L. (2016) The 3D Genome as Moderator of Chromosomal Communication.
2 *Cell* 164, 1110–1121

3 47 Duan, Z. *et al.* (2010) A three-dimensional model of the yeast genome. *Nature* 465, 363–367

4 48 Descorps-Declère, S. *et al.* (2015) Genome-wide replication landscape of *Candida glabrata*. *BMC*
5 *Biol.* 13, 69

6 49 Galazka, J.M. *et al.* (2016) *Neurospora* chromosomes are organized by blocks of importin alpha-
7 dependent heterochromatin that are largely independent of H3K9me3. *Genome Res.* 26, 1069–
8 1080

9 50 Kakui, Y. *et al.* (2017) Condensin-mediated remodeling of the mitotic chromatin landscape in
10 fission yeast. *Nat. Genet.* 49, 1553–1557

11 51 Marbouty, M. *et al.* (2014) Metagenomic chromosome conformation capture (meta3C) unveils
12 the diversity of chromosome organization in microorganisms. *eLife* 3, e03318

13 52 Mizuguchi, T. *et al.* (2014) Cohesin-dependent globules and heterochromatin shape 3D genome
14 architecture in *S. pombe*. *Nature* 516, 432–435

15 53 Marie-Nelly, H. *et al.* (2014) Filling annotation gaps in yeast genomes using genome-wide contact
16 maps. *Bioinforma. Oxf. Engl.* 30, 2105–2113

17 54 Varoquaux, N. *et al.* (2015) Accurate identification of centromere locations in yeast genomes
18 using Hi-C. *Nucleic Acids Res.* 43, 5331–5339

19 55 Burrack, L.S. *et al.* (2016) Neocentromeres Provide Chromosome Segregation Accuracy and
20 Centromere Clustering to Multiple Loci along a *Candida albicans* Chromosome. *PLoS Genet.* 12,
21 e1006317

22 56 Luo, J. *et al.* (2018) Karyotype engineering by chromosome fusion leads to reproductive isolation
23 in yeast. *Nature* 560, 392–396

24 57 Shao, Y. *et al.* (2018) Creating a functional single-chromosome yeast. *Nature* 560, 331–335

25 58 Cohen, B.A. *et al.* (2000) A computational analysis of whole-genome expression data reveals
26 chromosomal domains of gene expression. *Nat. Genet.* 26, 183–186

27 59 Janga, S.C. *et al.* (2008) Transcriptional regulation constrains the organization of genes on
28 eukaryotic chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15761–15766

29 60 Agier, N. *et al.* (2018) The evolution of the temporal program of genome replication. *Nat.*
30 *Commun.* 9, 2199

31 61 Raghuraman, M.K. *et al.* (2001) Replication dynamics of the yeast genome.

32 62 Vincenten, N. *et al.* (2015) The kinetochore prevents centromere-proximal crossover
33 recombination during meiosis. *eLife* 4,

34 63 Mascher, M. *et al.* (2017) A chromosome conformation capture ordered sequence of the barley
35 genome. *Nature* 544, 427–433

36 64 Dong, P. *et al.* (2017) 3D Chromatin Architecture of Large Plant Genomes Determined by Local
37 A/B Compartments. *Mol. Plant* 10, 1497–1509

38 65 Feng, S. *et al.* (2014) Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution
39 chromatin interactions in *Arabidopsis*. *Mol. Cell* 55, 694–707

40 66 Sexton, T. *et al.* (2012) Three-dimensional folding and functional organization principles of the
41 *Drosophila* genome. *Cell* 148, 458–472

42 67 Tolhuis, B. *et al.* (2011) Interactions among Polycomb domains are guided by chromosome
43 architecture. *PLoS Genet.* 7, e1001343

44 68 Zhang, Y. *et al.* (2012) Spatial organization of the mouse genome and its role in recurrent
45 chromosomal translocations. *Cell* 148, 908–921

46 69 Imakaev, M. *et al.* (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome
47 organization. *Nat. Methods* 9, 999–1003

48 70 Cournac, A. *et al.* (2016) The 3D folding of metazoan genomes correlates with the association of
49 similar repetitive elements. *Nucleic Acids Res.* 44, 245–255

50 71 Rao, S.S.P. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles
51 of chromatin looping. *Cell* 159, 1665–1680

1 72 Stevens, T.J. *et al.* (2017) 3D structures of individual mammalian genomes studied by single-cell
2 Hi-C. *Nature* 544, 59–64

3 73 Nishimura, K. *et al.* (2018) 3D genomic architecture reveals that neocentromeres associate with
4 heterochromatin regions. *J. Cell Biol.* DOI: 10.1083/jcb.201805003

5 74 Alonso, A. *et al.* (2010) A paucity of heterochromatin at functional human neocentromeres.
6 *Epigenetics Chromatin* 3, 6

7 75 Hori, T. *et al.* (2014) Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential
8 for kinetochore assembly. *Dev. Cell* 29, 740–749

9 76 Shang, W.-H. *et al.* (2013) Chromosome engineering allows the efficient isolation of vertebrate
10 neocentromeres. *Dev. Cell* 24, 635–648

11 77 Ay, F. *et al.* (2014) Three-dimensional modeling of the *P. falciparum* genome during the
12 erythrocytic cycle reveals a strong connection between genome architecture and gene
13 expression. *Genome Res.* 24, 974–988

14 78 Bunnik, E.M. *et al.* (2019) Comparative 3D genome organization in apicomplexan parasites. *Proc.*
15 *Natl. Acad. Sci. U. S. A.* 116, 3183–3192

16 79 Gibcus, J.H. *et al.* (2018) A pathway for mitotic chromosome formation. *Science* 359,
17 80 Lazar-Stefanita, L. *et al.* (2017) Cohesins and condensins orchestrate the 4D dynamics of yeast
18 chromosomes during the cell cycle. *EMBO J.* 36, 2684–2697

19 81 Naumova, N. *et al.* (2013) Organization of the mitotic chromosome. *Science* 342, 948–953

20 82 Grob, S. *et al.* (2013) Characterization of chromosomal architecture in *Arabidopsis* by
21 chromosome conformation capture. *Genome Biol.* 14, R129

22 83 Hou, C. *et al.* (2012) Gene density, transcription, and insulators contribute to the partition of the
23 *Drosophila* genome into physical domains. *Mol. Cell* 48, 471–484

24 84 Stadler, M.R. *et al.* (2017) Convergence of topological domain boundaries, insulators, and
25 polytene interbands revealed by high-resolution mapping of chromatin contacts in the early
26 *Drosophila melanogaster* embryo. *eLife* 6,
27 85 Kalhor, R. *et al.* (2011) Genome architectures revealed by tethered chromosome conformation
28 capture and population-based modeling. *Nat. Biotechnol.* 30, 90–98

29 86 Mercy, G. *et al.* (2017) 3D organization of synthetic and scrambled chromosomes. *Science* 355,
30 87 Therizols, P. *et al.* (2010) Chromosome arm length and nuclear constraints determine the
31 dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. U. S. A.* 107, 2025–2030

32 88 Jagannathan, M. *et al.* (2018) A conserved function for pericentromeric satellite DNA. *eLife* 7,
33 89 Nuebler, J. *et al.* (2018) Chromatin organization by an interplay of loop extrusion and
34 compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.* 115, E6697–E6706

35 90 Strom, A.R. *et al.* (2017) Phase separation drives heterochromatin domain formation. *Nature* 547,
36 241–245

37 91 Westermann, S. and Schleiffer, A. (2013) Family matters: structural and functional conservation
38 of centromere-associated proteins from yeast to humans. *Trends Cell Biol.* 23, 260–269

39 92 Woodruff, J.B. *et al.* (2018) Organization and Function of Non-dynamic Biomolecular
40 Condensates. *Trends Biochem. Sci.* 43, 81–94

41 93 Peng, J.C. and Karpen, G.H. (2007) H3K9 methylation and RNA interference regulate nucleolar
42 organization and repeated DNA stability. *Nat. Cell Biol.* 9, 25–35

43 94 Mekhail, K. *et al.* (2008) Role for perinuclear chromosome tethering in maintenance of genome
44 stability. *Nature* 456, 667–670

45 95 Simon, L. *et al.* (2015) Structure and Function of Centromeric and Pericentromeric
46 Heterochromatin in *Arabidopsis thaliana*. *Front. Plant Sci.* 6, 1049

47 96 Martin, C. *et al.* (2006) Genome restructuring in mouse embryos during reprogramming and early
48 development. *Dev. Biol.* 292, 317–332

49 97 Ioannou, D. *et al.* (2017) A new model of sperm nuclear architecture following assessment of the
50 organization of centromeres and telomeres in three-dimensions. *Sci. Rep.* 7, 41585

- 1 98 Tjong, H. *et al.* (2016) Population-based 3D genome structure analysis reveals driving forces in
2 spatial genome organization. *Proc. Natl. Acad. Sci. U. S. A.* 113, E1663-1672
- 3 99 Fitzgerald-Hayes, M. *et al.* (1982) Nucleotide sequence comparisons and functional analysis of
4 yeast centromere DNAs. *Cell* 29, 235–244
- 5 100 Furuyama, S. and Biggins, S. (2007) Centromere identity is specified by a single centromeric
6 nucleosome in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14706–14711
- 7 101 Wood, V. *et al.* (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415, 871–
8 880
- 9 102 Chatterjee, G. *et al.* (2016) Repeat-Associated Fission Yeast-Like Regional Centromeres in the
10 Ascomycetous Budding Yeast *Candida tropicalis*. *PLoS Genet.* 12, e1005839
- 11 103 Yadav, V. *et al.* (2018) RNAi is a critical determinant of centromere evolution in closely related
12 fungi. *Proc. Natl. Acad. Sci. U. S. A.* 115, 3108–3113
- 13 104 Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant
14 *Arabidopsis thaliana*. *Nature* 408, 796–815
- 15 105 Choulet, F. *et al.* (2014) Structural and functional partitioning of bread wheat chromosome 3B.
16 *Science* 345, 1249721
- 17 106 International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based
18 draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345, 1251788
- 19 107 Schnable, P.S. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*
20 326, 1112–1115
- 21 108 Wolfgruber, T.K. *et al.* (2009) Maize centromere structure and evolution: sequence analysis of
22 centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. *PLoS Genet.* 5,
23 e1000743
- 24 109 Talbert, P.B. *et al.* (2018) Simple and Complex Centromeric Satellites in *Drosophila* Sibling
25 Species. *Genetics* 208, 977–990
- 26 110 Komissarov, A.S. *et al.* (2011) Tandemly repeated DNA families in the mouse genome. *BMC*
27 *Genomics* 12, 531
- 28 111 Kuznetsova, I. *et al.* (2006) High-resolution organization of mouse centromeric and
29 pericentromeric DNA. *Cytogenet. Genome Res.* 112, 248–255
- 30 112 Aldrup-Macdonald, M.E. and Sullivan, B.A. (2014) The past, present, and future of human
31 centromere genomics. *Genes* 5, 33–50
- 32 113 Henikoff, J.G. *et al.* (2015) A unique chromatin complex occupies young α -satellite arrays of
33 human centromeres. *Sci. Adv.* 1,
- 34 114 Markaki, Y. *et al.* (2013) Fluorescence in situ hybridization applications for super-resolution 3D
35 structured illumination microscopy. *Methods Mol. Biol. Clifton NJ* 950, 43–64
- 36 115 Chen, Y. *et al.* (2000) The N terminus of the centromere H3-like protein Cse4p performs an
37 essential function distinct from that of the histone fold domain. *Mol. Cell. Biol.* 20, 7037–7048
- 38 116 Servant, N. *et al.* (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing.
39 *Genome Biol.* 16, 259
- 40 117 Wolff, J. *et al.* (2018) Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality
41 control and visualization. *Nucleic Acids Res.* 46, W11–W16
- 42 118 Servant, N. *et al.* (2012) HiTC: exploration of high-throughput “C” experiments. *Bioinforma. Oxf.*
43 *Engl.* 28, 2843–2844
- 44
- 45

46